

PROCEEDINGS
NINTH ANNUAL ALLERTON CONFERENCE
ON CIRCUIT AND SYSTEM THEORY

J. B. Cruz, Jr.
A. H. Haddad

Conference Co-Chairmen

Conference held
October 6, 7, 8, 1971
Allerton House
Monticello, Illinois

Sponsored by the
DEPARTMENT OF ELECTRICAL ENGINEERING
and the
COORDINATED SCIENCE LABORATORY
of the
UNIVERSITY OF ILLINOIS
at
Urbana-Champaign

J.W. BANDLER and C. CHARALAMBOUS
 Communications Research Laboratory and
 Department of Electrical Engineering
 McMaster University
 Hamilton, Ontario, Canada

Abstract This paper presents a theoretical discussion of the necessary and sufficient conditions for optimality in generalized nonlinear least p th approximation problems for $p \rightarrow \infty$. In the limit the conditions for a minimax approximation are derived, as is to be expected. Numerical examples involving the modelling of a fourth-order system by a second-order model, and quarter-wave transmission-line transformers illustrate the results.

INTRODUCTION

Of great practical importance to network and system designers wishing to approximate a specified response by a network or system response, or desiring to meet or exceed certain design specifications is the optimality of their approximation. A number of workers interested in minimax approximation [1,2,3] have independently arrived at similar conditions for nonlinear minimax approximation problems. These are naturally derivable from the Kuhn-Tucker conditions for a constrained optimum because of the close relationship between nonlinear minimax approximation and nonlinear programming. Bandler [3], in particular, derived the appropriate conditions in a general form suitable for such problems as filter design.

Because of the widespread interest in nonlinear least p th approximation [4], and because of recent results [5] that permit least p th objectives in a more generalized sense to be directly applicable to such problems as meeting or exceeding design specifications as in filter design it is felt that a detailed mathematical discussion of conditions for optimality is highly relevant. Thus, the present paper allows for situations more general than the conventional problem of approximation to a single continuous function on a closed interval.

DEFINITIONS AND ASSUMPTIONS

Define real error functions related to the "upper" and "lower" specifications, respectively, as [5], [6]

$$\begin{aligned} e_u(\phi, \psi) &\triangleq w_u(\psi) (F(\phi, \psi) - S_u(\psi)) \\ e_l(\phi, \psi) &\triangleq w_l(\psi) (F(\phi, \psi) - S_l(\psi)) \end{aligned} \quad (1)$$

where

- $F(\phi, \psi)$ is the approximating function (actual response)
- $S_u(\psi)$ is an upper specified function (desired response bound)
- $S_l(\psi)$ is a lower specified function (desired response bound)
- $w_u(\psi)$ is an upper positive weighting function
- $w_l(\psi)$ is a lower positive weighting function
- ϕ is a vector containing the k independent parameters
- ψ is an independent variable (e.g., frequency or time)

This work was supported by the National Research Council of Canada under grant A7239 and by a Frederick Gardner Cottrell grant from the Research Corporation.

In practice, we will evaluate all the functions at a finite discrete set of values of ψ taken from one or more closed intervals. Therefore, we will let

$$\begin{aligned} e_{ui}(\phi) &\triangleq e_u(\phi, \psi_i) & i \in I_u \\ e_{li}(\phi) &\triangleq e_l(\phi, \psi_i) & i \in I_l \end{aligned} \quad (2)$$

where it is assumed that a sufficient number of sample points have been chosen so that the discrete approximation problem adequately approximates the continuous problem. I_u and I_l are appropriate index sets.

Case 1 - Specification Violated In the case when the specification is violated some of the $e_{ui}(\phi)$ or $-e_{li}(\phi)$ are positive. In an effort to meet the specification we can propose the following objective function to be minimized¹

$$U(\phi) = \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_l} [-e_{li}(\phi)]^p \right\}^{\frac{1}{p}} \quad (3)$$

where

$$\begin{aligned} J_u &\triangleq \{i | e_{ui}(\phi) \geq 0, \quad i \in I_u\} \\ J_l &\triangleq \{i | -e_{li}(\phi) \geq 0, \quad i \in I_l\} \end{aligned} \quad (4)$$

and $p > 1$.

The larger the value of p the more nearly would we expect the maximum error to be emphasized, since

$$\max_i [e_{ui}(\phi), -e_{li}(\phi)] = \lim_{p \rightarrow \infty} \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_l} [-e_{li}(\phi)]^p \right\}^{\frac{1}{p}}. \quad (5)$$

Case 2 - Specification Satisfied For the case when the specification is satisfied all the $-e_{ui}(\phi)$ and $e_{li}(\phi)$ will be positive. This time, in an effort to exceed the specification by as much as possible, we can propose the following objective function to be minimized¹

$$U(\phi) = - \left\{ \sum_{i \in I_u} [-e_{ui}(\phi)]^{-p} + \sum_{i \in I_l} [e_{li}(\phi)]^{-p} \right\}^{-\frac{1}{p}} \quad (6)$$

where we assume

$$\begin{aligned} -e_{ui}(\phi) &> 0 & i \in I_u \\ e_{li}(\phi) &> 0 & i \in I_l \end{aligned} \quad (7)$$

and $p \geq 1$.

The larger the value of p the more nearly would we expect the minimum "error" to be emphasized, since

$$\min_i [-e_{ui}(\phi), e_{li}(\phi)] = \lim_{p \rightarrow \infty} \left\{ \sum_{i \in I_u} [-e_{ui}(\phi)]^{-p} + \sum_{i \in I_l} [e_{li}(\phi)]^{-p} \right\}^{-\frac{1}{p}}. \quad (8)$$

Assumptions It is assumed that a minimum exists in a closed and bounded region of points ϕ and that $e_{ui}(\phi)$ and $e_{li}(\phi)$ are continuous for all i with continuous partial derivatives at least in the neighborhood of the minimum.

¹Since, in this paper, the conditions for optimality are of interest, it is convenient to use slightly different objective functions from those proposed in an earlier work [5].

Then the objective functions proposed are continuous with continuous partial derivatives in the neighborhood of the minimum.

THEOREMS

Theorem 1 At an optimum point $\check{\phi}_\infty$ for a minimax approximation problem

$$\sum_{i \in K_u} u_{ui} \nabla e_{ui}(\check{\phi}_\infty) = \sum_{i \in K_\ell} u_{\ell i} \nabla e_{\ell i}(\check{\phi}_\infty)$$

$$\sum_{i \in K_u} u_{ui} + \sum_{i \in K_\ell} u_{\ell i} = 1$$

$$u_{ui} \geq 0 \quad i \in K_u$$

$$u_{\ell i} \geq 0 \quad i \in K_\ell$$

where

$$\nabla \Delta = \left[\frac{\partial}{\partial \phi_1} \quad \frac{\partial}{\partial \phi_2} \quad \dots \quad \frac{\partial}{\partial \phi_k} \right]^T$$

and where $e_{ui}(\check{\phi}_\infty)$ for $i \in K_u$ and $-e_{\ell i}(\check{\phi}_\infty)$ for $i \in K_\ell$ are the equal maxima.

Proof for Case 1 Differentiating equation (3)

$$\begin{aligned} \nabla U(\phi) &= \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi)]^p \right\}^{\frac{1}{p}-1} \\ &\quad \cdot \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^{p-1} \nabla e_{ui}(\phi) - \sum_{i \in J_\ell} [-e_{\ell i}(\phi)]^{p-1} \nabla e_{\ell i}(\phi) \right\} \\ &= \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi)]^p \right\}^{\frac{1}{p}} \\ &\quad \cdot \left\{ \sum_{i \in J_u} \frac{[e_{ui}(\phi)]^p}{\sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi)]^p} \frac{\nabla e_{ui}(\phi)}{e_{ui}(\phi)} \right. \\ &\quad \left. - \sum_{i \in J_\ell} \frac{[-e_{\ell i}(\phi)]^p}{\sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi)]^p} \frac{\nabla e_{\ell i}(\phi)}{[-e_{\ell i}(\phi)]} \right\} \end{aligned} \quad (9)$$

The necessary conditions for an optimum of $U(\phi)$ are that

$$\nabla U(\check{\phi}_p) = 0 \quad (10)$$

where $\check{\phi}_p$ denotes the optimum parameter vector for particular values of p .
Let

$$\mu_i(p) = \frac{[e_{ui}(\check{\phi}_p)]^p}{\sum_{i \in J_u} [e_{ui}(\check{\phi}_p)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\check{\phi}_p)]^p} \quad (11)$$

$$\lambda_i(p) = \frac{[-e_{\ell i}(\check{\phi}_p)]^p}{\sum_{i \in J_u} [e_{ui}(\check{\phi}_p)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\check{\phi}_p)]^p} \quad (12)$$

$$u_{ui} = \lim_{p \rightarrow \infty} u_i(p) \quad (13)$$

$$u_{\ell i} = \lim_{p \rightarrow \infty} \lambda_i(p) \quad (14)$$

Then it is clear that as $p \rightarrow \infty$, the necessary conditions for optimality applied to equation (9) yield

$$u_{ui} \begin{cases} = 0 & i \notin K_u \\ \geq 0 & i \in K_u \end{cases} \quad (15)$$

$$u_{\ell i} \begin{cases} = 0 & i \notin K_\ell \\ \geq 0 & i \in K_\ell \end{cases} \quad (16)$$

and

$$\sum_{i \in K_u} u_{ui} + \sum_{i \in K_\ell} u_{\ell i} = 1 \quad (17)$$

Therefore, as $p \rightarrow \infty$,

$$\sum_{i \in K_u} u_{ui} \nabla e_{ui}(\check{\phi}_\infty) = \sum_{i \in K_\ell} u_{\ell i} \nabla e_{\ell i}(\check{\phi}_\infty) \quad (18)$$

and the theorem is proved.

Proof for Case 2 Following a similar procedure to the one used for Case 1, the same conditions can be derived but in this case we let

$$u_i(p) = \frac{[-e_{ui}(\check{\phi}_p)]^{-p}}{\sum_{i \in I_u} [-e_{ui}(\check{\phi}_p)]^{-p} + \sum_{i \in I_\ell} [e_{\ell i}(\check{\phi}_p)]^{-p}} \quad (19)$$

$$\lambda_i(p) = \frac{[e_{\ell i}(\check{\phi}_p)]^{-p}}{\sum_{i \in I_u} [-e_{ui}(\check{\phi}_p)]^{-p} + \sum_{i \in I_\ell} [e_{\ell i}(\check{\phi}_p)]^{-p}} \quad (20)$$

Theorem 2 If the relations in Theorem 1 are satisfied at a point $\check{\phi}_\infty$ and the $e_{ui}(\check{\phi})$ are convex for $i \in I_u$ and the $e_{\ell i}(\check{\phi})$ are concave for $i \in I_\ell$ then $\check{\phi}_\infty$ is optimal.

Proof for Case 1 Assuming the $e_{ui}(\check{\phi})$ are convex and the $-e_{\ell i}(\check{\phi})$ are convex

$$\begin{aligned} e_{ui}((1-\lambda)\phi^1 + \lambda\phi^2) &\leq (1-\lambda)e_{ui}(\phi^1) + \lambda e_{ui}(\phi^2) \\ -e_{\ell i}((1-\lambda)\phi^1 + \lambda\phi^2) &\leq -(1-\lambda)e_{\ell i}(\phi^1) - \lambda e_{\ell i}(\phi^2) \end{aligned} \quad (21)$$

for $0 \leq \lambda \leq 1$. Therefore, for $p \geq 1$,

$$\begin{aligned} &\left\{ \sum_{i \in J_u} [e_{ui}((1-\lambda)\phi^1 + \lambda\phi^2)]^p + \sum_{i \in J_\ell} [-e_{\ell i}((1-\lambda)\phi^1 + \lambda\phi^2)]^p \right\}^{\frac{1}{p}} \\ &\leq \left\{ \sum_{i \in J_u} [(1-\lambda)e_{ui}(\phi^1) + \lambda e_{ui}(\phi^2)]^p \right. \\ &\quad \left. + \sum_{i \in J_\ell} [-(1-\lambda)e_{\ell i}(\phi^1) - \lambda e_{\ell i}(\phi^2)]^p \right\}^{\frac{1}{p}} \end{aligned}$$

$$\begin{aligned} &\leq (1-\lambda) \left\{ \sum_{i \in J_u} [e_{ui}(\phi^1)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi^1)]^p \right\}^{\frac{1}{p}} \\ &+ \lambda \left\{ \sum_{i \in J_u} [e_{ui}(\phi^2)]^p + \sum_{i \in J_\ell} [-e_{\ell i}(\phi^2)]^p \right\}^{\frac{1}{p}} \end{aligned} \quad (22)$$

from Minkowski's inequality [7]. Hence,

$$U((1-\lambda)\phi^1 + \lambda\phi^2) \leq (1-\lambda)U(\phi^1) + \lambda U(\phi^2) \quad (23)$$

Therefore, $U(\phi)$ is convex and Theorem 2 follows.

Proof for Case 2 Using equation (21) we have, for $p \geq 1$,

$$\begin{aligned} &\left\{ \sum_{i \in I_u} [-e_{ui}((1-\lambda)\phi^1 + \lambda\phi^2)]^{-p} + \sum_{i \in I_\ell} [e_{\ell i}((1-\lambda)\phi^1 + \lambda\phi^2)]^{-p} \right\}^{-\frac{1}{p}} \\ &\geq \left\{ \sum_{i \in I_u} [-(1-\lambda)e_{ui}(\phi^1) - \lambda e_{ui}(\phi^2)]^{-p} \right. \\ &+ \left. \sum_{i \in I_\ell} [(1-\lambda)e_{\ell i}(\phi^1) + \lambda e_{\ell i}(\phi^2)]^{-p} \right\}^{-\frac{1}{p}} \\ &\geq (1-\lambda) \left\{ \sum_{i \in I_u} [-e_{ui}(\phi^1)]^{-p} + \sum_{i \in I_\ell} [e_{\ell i}(\phi^1)]^{-p} \right\}^{-\frac{1}{p}} \\ &+ \lambda \left\{ \sum_{i \in I_u} [-e_{ui}(\phi^2)]^{-p} + \sum_{i \in I_\ell} [e_{\ell i}(\phi^2)]^{-p} \right\}^{-\frac{1}{p}} \end{aligned} \quad (24)$$

Space does not permit a proof of this inequality, which is a counterpart to Minkowski's inequality. Hence,

$$-U((1-\lambda)\phi^1 + \lambda\phi^2) \geq -(1-\lambda)U(\phi^1) - \lambda U(\phi^2) \quad (25)$$

Therefore, $U(\phi)$ is convex and Theorem 2 follows.

EXAMPLES

Example 1 The first example is to find a second-order model of a fourth-order system when the input to the system is an impulse for different values of p . The transfer function of the fourth-order system is

$$G(s) = \frac{\lambda(s+4)}{(s+1)(s^2+4s+8)(s+5)} \quad (26)$$

The transfer function of the second-order model considered is

$$H(s) = \frac{c}{(s+\alpha)^2 + \beta^2} \quad (27)$$

Therefore, in our case we have

$$S(t) = \frac{3}{20} \exp(-t) + \frac{1}{52} \exp(-5t) - \frac{\exp(-2t)}{65} (3\sin 2t + 11\cos 2t) \quad (28)$$

$$F(\phi, t) = \frac{c}{\beta} \exp(-\alpha t) \sin \beta t \quad (29)$$

$$\phi = [\alpha \quad \beta \quad c]^T \quad (30)$$

$$\psi_i = \tau_i \quad 0 \leq \tau_i \leq 20 \quad (31)$$

$$e(\phi, \tau) = F(\phi, \psi) - S(\psi) \quad (32)$$

$$U(\phi) = \left\{ \sum_{i=1}^{101} |e(\phi, \tau_i)|^p \right\}^{\frac{1}{p}} \quad (33)$$

The above objective function, which corresponds to the situation $S_U=S_L=S$ and $w_U=w_L=1$ in Case 1, was set up using 101 uniformly spaced points over the range 0 and 20 sec. The values of p used for optimization² were 2, 6, 10, 20, 30, 40, 50, 70, 80, 100, 200, 1000 and 10,000. See Figure 1.

The values of $\mu_i(p)$ and $\lambda_i(p)$ were calculated for different values of p and for different values of t . Seven values of t were considered, of which four were the points where the approximately equal extrema occurred. Figure 2 demonstrates the validity of equations (15) and (16).

Example 2 Here, we consider the design of 10Ω to 1Ω quarter-wave transmission-line transformers taking, for convenience, 0.3 as an upper specification for the magnitude of the reflection coefficient ρ over a specified 100% frequency band. The basic problem has been previously defined and analyzed in the context of optimality by Bandler [3]. See also [9] and [10]. The optimum 2-section transformer violates the specification. The optimum 3-section transformer, on the other hand, satisfies the specification. The value of p for both cases in the optimization process² was 10,000. 101 uniformly spaced sample points were used.

The maximum values of $e_{U1}(\phi_p)$, where $F = |\rho|$ and $S_U = 0.3$, and the frequencies at which they occur are shown in Tables I and II. Using equation (11) and Table I non-zero multipliers in the ratio

$$3.001 : 2.001 : 1.000$$

and using equation (19) and Table II non-zero multipliers in the ratio

$$2.999 : 2.807 : 1.759 : 1.000$$

can be found. These results may be compared with those of Bandler [3].

CONCLUSIONS

It has been shown [5] that, for violated specifications in generalized least p th approximation, we have a close analogy with minimization of a penalty term designed to bring one closer to the boundary of the feasible region, whereas, for satisfied specifications, we have an analogy with minimization of a penalty term designed to steer a feasible solution deeper into the feasible region. Conventional approximation problems fit into the former category, but conditions of optimality are required for both.

In the limit the conditions for a minimax approximation are obtained as is to be expected. It is felt that new insight into minimax algorithms is gained, since appropriate algorithms might attempt to force these conditions in an iterative manner by using extremely large values of p .

²Techniques for least p th approximation with extremely large values of p are described elsewhere [8].

TABLE I
 OPTIMUM 2-SECTION TRANSFORMER WITH $p=10^4$
 (SPECIFICATION VIOLATED - CASE 1)

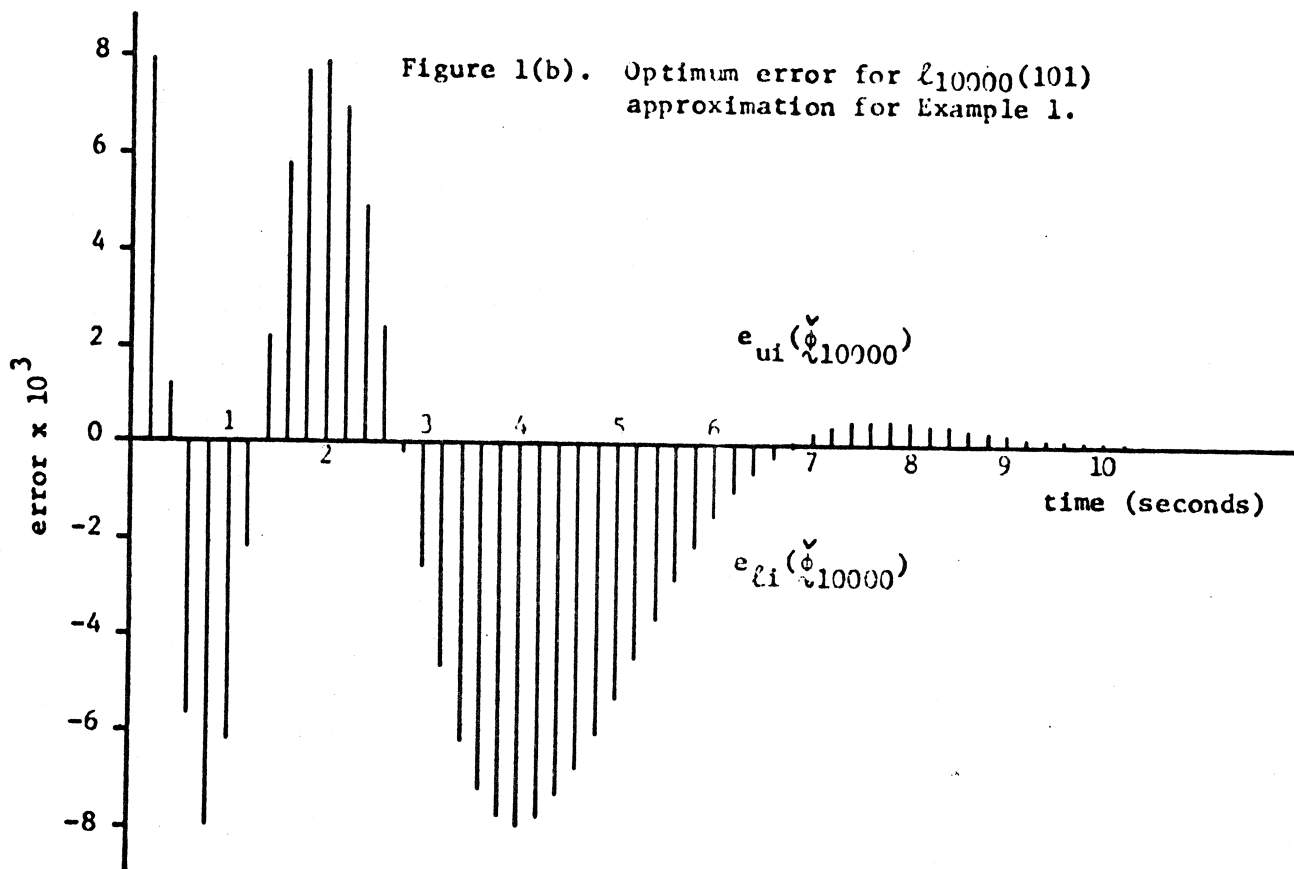
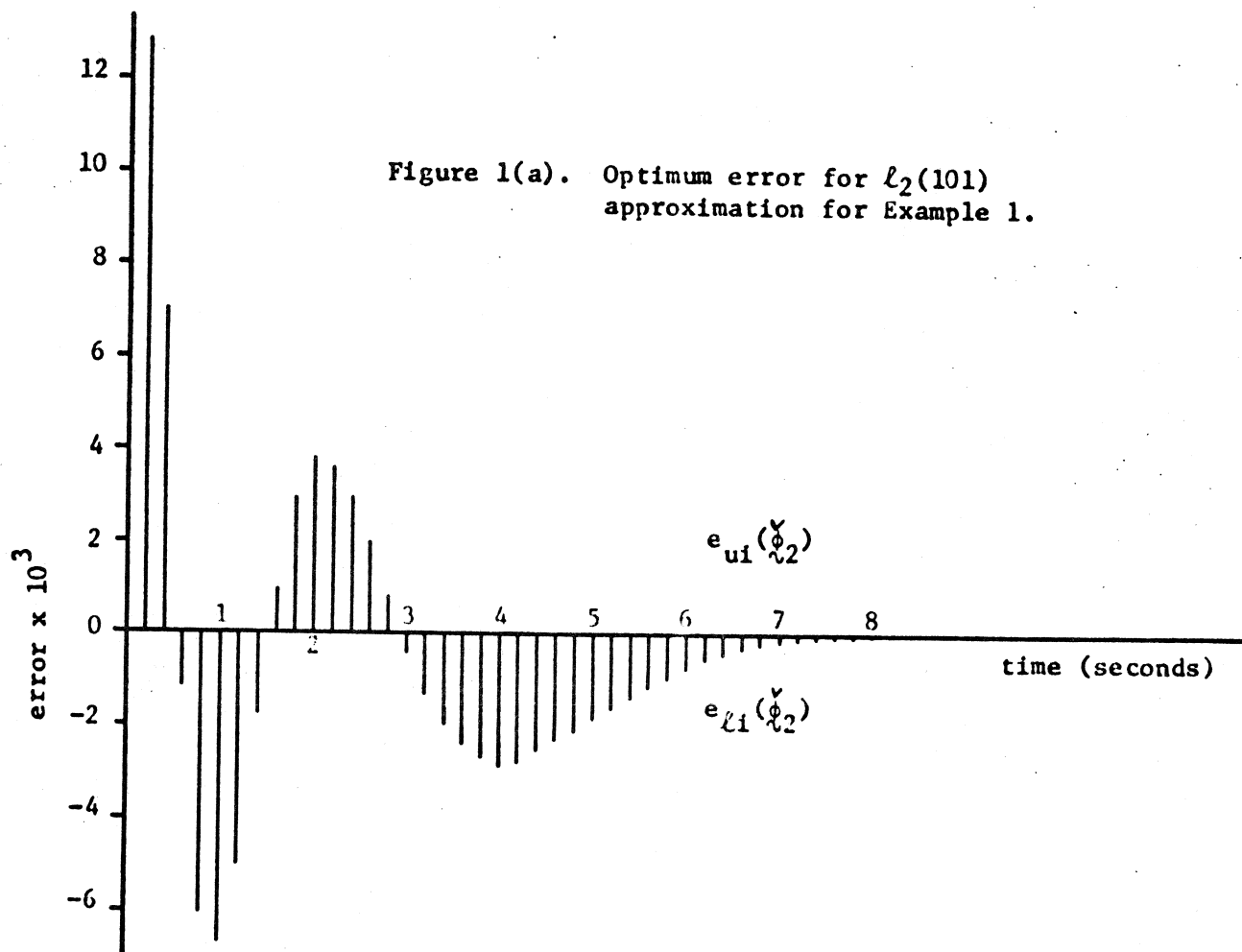
Maximum of $e_{ui}(\frac{\gamma}{\lambda_p})$	0.12857552	0.12857031	0.12856139
Frequencies (GHz) at which maximum occurs	0.5	1.0	1.5

TABLE II
 OPTIMUM 3-SECTION TRANSFORMER WITH $p=10^4$
 (SPECIFICATION SATISFIED - CASE 2)

Maximum of $-e_{ui}(\frac{\gamma}{\lambda_p})$	0.10270671	0.10270739	0.10271219	0.10271799
Frequencies (GHz) at which maximum occurs	0.5	0.77	1.23	1.5

REFERENCES

- [1] V.F. Dem'yanov, "Sufficient conditions for minimax problems", Zh. Uychisl. Mat. Fiz., vol. 10, pp. 1107-1115, 1970.
- [2] J. Medanic, "Solution of the convex minimax problem by the Newton-Raphson method", Proc. 8th Annual Allerton Conf. on Circuit and System Theory, Urbana, Ill., pp. 13-22, October 1970.
- [3] J.W. Bandler, "Conditions for a minimax optimum", Proc. 8th Annual Allerton Conf. on Circuit and System Theory, Urbana, Ill., pp. 23-30, October 1970. See also IEEE Trans. Circuit Theory, vol. CT-18, pp. 476-479, July 1971.
- [4] G.C. Temes and D.Y.F. Zai, "Least pth approximation", IEEE Trans. Circuit Theory, vol. CT-16, pp. 235-237, May 1969.
- [5] J.W. Bandler and C. Charalambous, "Generalized least pth objectives for networks and systems", Proc. 14th Midwest Symposium on Circuit Theory, Denver, Colo., pp. 10.5.1-10.5.10, May 1971.
- [6] J.W. Bandler, "Optimization methods for computer-aided design", IEEE Trans. Microwave Theory and Techniques, vol. MTT-17, pp. 533-552, August 1969.
- [7] N.I. Achieser, Theory of Approximation (Transl. by C.J. Hyman). New York: Frederick Unger, 1956.
- [8] J.W. Bandler and C. Charalambous, "Practical least pth approximation with extremely large values of p", 5th Asilomar Conf. on Circuits and Systems, Pacific Grove, Calif., November 1971.
- [9] J.W. Bandler and P.A. Macdonald, "Cascaded noncommensurate transmission-line networks as optimization problems", IEEE Trans. Circuit Theory, vol. CT-16, pp. 391-394, August 1969.
- [10] J.W. Bandler and P.A. Macdonald, "Optimization of microwave networks by razor search", IEEE Trans. Microwave Theory and Techniques, vol. MTT-17, pp. 552-562, August 1969.



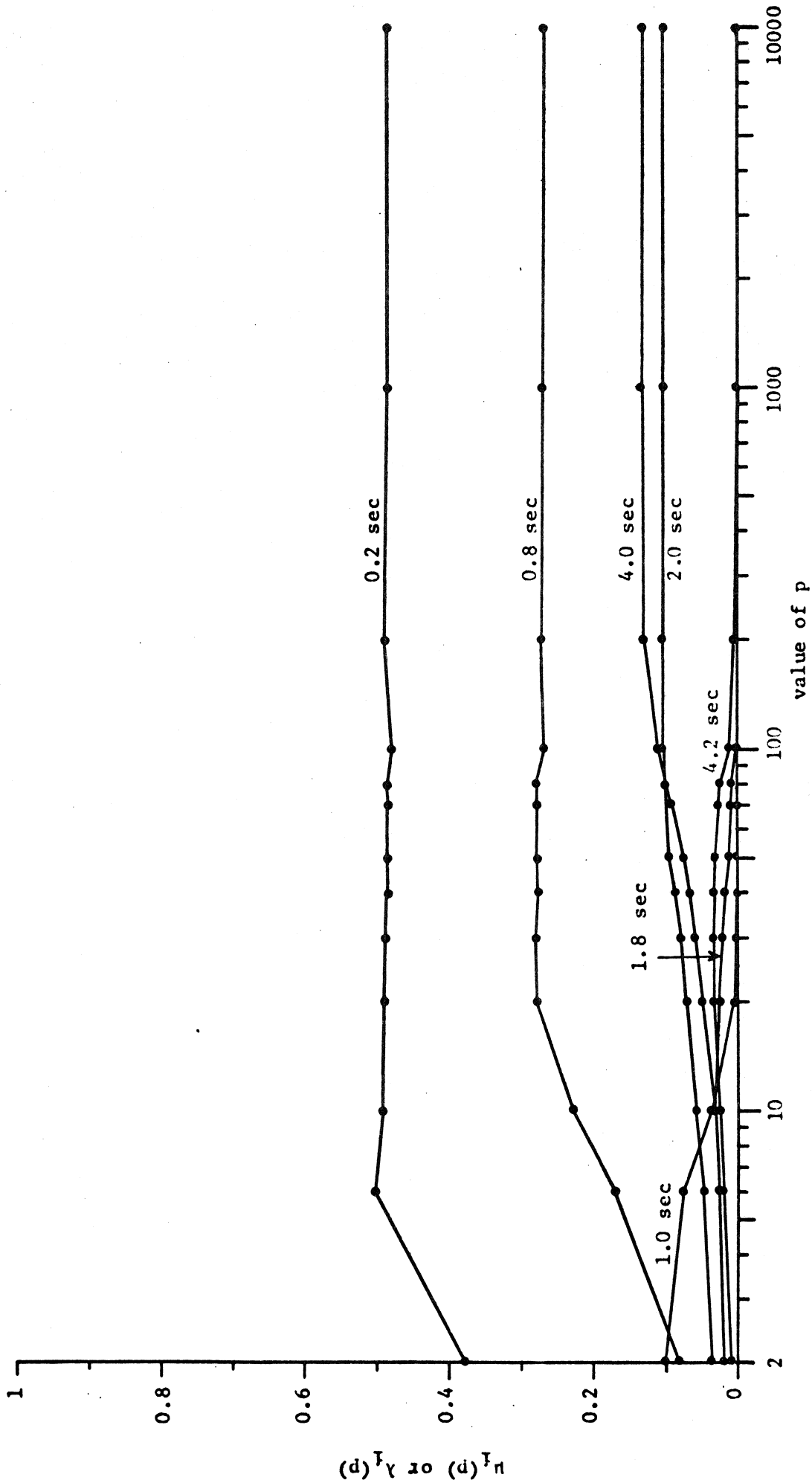


Figure 2. $\mu_i(p)$ or $\lambda_i(p)$, as appropriate, calculated at specified values of time and for certain values of p for Example 1.